Estimation of influenza epidemic model parameters from reported mortality data*

Eugene M. Gavrilov
Computer and Computational Sciences Division,
Los Alamos National Laboratory,
MS B287 Los Alamos New Mexico 87544, *gavrilov@lanl.gov*

Alan S. Perelson
Theoretical Division, Los Alamos National Laboratory,
MS K710, Los Alamos New Mexico 87544, *asp@lanl.gov*

Alan Lapedes
Theoretical Division, Los Alamos National Laboratory,
MS B213, Los Alamos New Mexico 87544, *asl@lanl.gov*

Ira M. Longini, Jr.
Department of Biostatistics, Rollins School of Public Health,
Emory University, Atlanta Georgia 30322, *Longini@sph.emory.edu*

*LAUR-02-6467, Submitted to *Mathematical Biosciences*, September 2002

Abstract

We validate the ability of the simple SIR model to describe the spread of influenza for large metropolitan areas in the US. We derive an evolutionary equation for the excess death rates due to influenza and estimate parameters in the equation from the US weekly mortality/morbidity reports using the Random Collocation Least Squares method.
keywords: influenza, parameter estimation, least-squares, SIR model.

1 Introduction

Influenza is a major public health threat. In the U.S. influenza epidemics occur during the winter months between November and April and are responsible for an average of approximately 20,000 deaths/year [1]- [6]. Weekly influenza and pneumonia mortality data provide the major national source of information for monitoring the timing and severity of the yearly influenza epidemic in the U.S. [4]. Thus, it is of importance to extract basic epidemiologic information about yearly influenza epidemics from the reported mortality data. However, such mortality data are a mixture of deaths from influenza and from other causes. In this paper, we present a method for statistically separating influenza related mortality from other source of pneumonia mortality. In addition, our methods, based on the classical SIR epidemic model, are used to estimate important epidemiologic parameters.

In section 2, we present the mathematical form of the model. In section 3, we give the numerical form of the model that is fit to the mortality data. Section 4, gives details on the method of random collocation that we use to estimate parameters. Results are presented in section 5, followed by a discussion in section 6.

2 The SIR model

Consider the SIR model applied to the confined single population group. The infectious agent is spread through population from infected individuals to susceptibles. At any given moment the population can be roughly clustered into three distinct groups, i.e., $S(t)$ susceptible, $I(t)$ infected, and $R(t)$ recovered. We will assume random mixing among those population compartments. The assumption of a uniform spatial population distribution allows us to employ the concept of random mixing of all population groups. The recovered group is assumed to acquire immunity and does not interact with either of the other groups. To study the evolution of the disease spread we

then have the following nonlinear model:

$$\begin{aligned} \dot{S} &= -\alpha_1 S(I/N) \\ \dot{I} &= \alpha_1 I(S/N) - (\alpha_2 + \alpha_3) I \\ \dot{R} &= \alpha_2 I \end{aligned} \tag{1}$$

Here, α_1 is the infection rate, α_2 is the recovery rate, α_3 is the death rate due to the infection and $N(t)$ is the total population size $N(t) = (S(t) + R(t) + I(t))$,

$$\dot{N} = -\alpha_3 I. \tag{2}$$

We use dot for time derivative, i.e., $\dot{S} \equiv dS/dt$. The basic reproductive number R_0 is defined as $R_0 = \alpha_1/\alpha_2$. The system (1) states that the major cause of population decline is due to the death of infected people. This model is simplified compared to models of this type that usually include birth rate and death rates due to normal causes, as well as loss of immunity and other factors. However, for the study of severe pandemic disease spread when characteristic times are of the order of days and death rates are very high compare to the death rate due to normal causes, the approximation is justified. In addition, during normal flu seasons total deaths by other causes roughly equals total births, and we assume the number of susceptibles introduced by birth is low enough to be ignored.

There are two goals we would like to achieve. First, we wish to prove the concept that the model can correctly describe the spread of the infection not only qualitatively but quantitatively, provided the assumptions mentioned above hold true. The second goal is to estimate parameters and analyse the relation of the parameters to other macroscopic quantities like the density of the population, demographic and transportation specifics. There are arguments in the literature stating that the model (1) can be generalized for the multi-season case, if one introduces time dependency for the interaction coefficients. One of the arguments to support this is the fact that excess death rates for influenza have consistent repeating patterns over many years, see Fig. (1). The assumption being made is that the probability of interaction has a time component due to some fundamental but unknown reason. However even today there is no clear explanation for that reason and time structure for interaction is introduced *ad hoc*. We will concentrate our study on single-season outbreaks. We will estimate the parameters for the model employing the Random Collocation Least Squares (RCLS) technique. We will exploit a real data set acquired from the US National Center of Health Statistics (NCHS) and also the data provided to us by the Influenza Branch of the Centers for Disease Control and Prevention (CDC).

3 Numerical model and data

The data set we analyze represents the set of historic time-series of excess death rates due to pneumonia and influenza, reported by the NCHS on the weekly basis (MMWR weekly reports). The data is for 122 major US metropolitan areas and the total time-series is for a period of about four years for each of the cities. There are errors in this data set due to the way the data is collected and reported. The lack of direct information about the actual reported cases of infected people as well as those recovered (removed) does not allow one to use the model (1) directly for parameter estimation. However, we can exploit the structure of the model to derive an equation governing the evolution of the death rates. To proceed, we reduce the system (1) to a single equation by substituting $N - R - I$ for S, where

$$R(t) = R(t_0) + \int_{t_0}^{t} \alpha_2 I dt \tag{3}$$

and

$$N(t) = N(t_0) - \int_{t_0}^{t} \alpha_3 I dt \tag{4}$$

This yields

$$\dot{I} = (\alpha_1/N) \left[N(t_0) - \int_{t_0}^{t} \alpha_3 I dt - (R(t_0) + \int_{t_0}^{t} \alpha_2 I dt) - I \right] I - (\alpha_2 + \alpha_3) I \tag{5}$$

where t_0 is some arbitrary time before t. In the original system (1) the term $\alpha_3 I$ represents the influenza death rate. Thus, we can rewrite equation (5) in terms of the death-rate $D(t) = \alpha_3 I(t)$. Let us also expand N in denominator and keep all terms up to the order $O(\delta N)$. This means we will keep $N(t_0) + \delta N$ in the numerator and $N(t_0)$ in the denominator.

$$\begin{aligned} \dot{D} = \left[\tfrac{\alpha_1}{N(t_0)} (N(t_0) - R(t_0)) - (\alpha_2 + \alpha_3) \right] D - \\ \tfrac{\alpha_1(\alpha_2+\alpha_3)}{\alpha_3 N(t_0)} D \int_{t_0}^{t} D dt - \tfrac{\alpha_1}{\alpha_3 N(t_0)} D^2 \end{aligned} \tag{6}$$

From (6) one can see that just from one time-series for $D(t)$ it is possible to estimate all of the parameters α_i. Of those parameters α_2, the recovery rate is fairly well known, and the inverse of α_2 is close to one week (see [12]). We will use this information and fix this parameter to the value $\alpha_2 = 1.3$ for weekly sampled data. We will assume $R(t_0) = 0$ at the beginning of the season t_0 and put $N(t_0) = N_0$. In a typical flu season (16 weeks) in the US about 60 million people are infected and about 20,000 of these people die of

influenza [5] - [6]. Thus, α_3 is on the order of $\sim 2 \times 10^{-5}$. Clearly, $\alpha_2 \gg \alpha_3$ and we can use α_2 instead of $\alpha_2 + \alpha_3$ for the second term in front of the integral, however we will keep α_3 in the first term because α_1 can be close to α_2. Equation (6) then becomes

$$\dot{D} = (\alpha_1 - \alpha_2 - \alpha_3)\, D - \frac{\alpha_1}{N_0 \alpha_3}(\alpha_2 D \int_{t_0}^{t} D dt + D^2), \tag{7}$$

or equivalently

$$\dot{D} = \xi_1 D - \xi_2(\alpha_2 D \int_{t_0}^{t} D dt + D^2), \tag{8}$$

where

$$\begin{array}{l} \xi_1 = \alpha_1 - \alpha_2 - \alpha_3 \\ \xi_2 = \alpha_1/(\alpha_3 N_0). \end{array} \tag{9}$$

The parameters α_1, α_3 are

$$\begin{array}{l} \alpha_1 = N_0 \xi_2 (\xi_1 + \alpha_2)/(N_0 \xi_2 - 1) \\ \alpha_3 = (\xi_1 + \alpha_2)/(N_0 \xi_2 - 1) \end{array} \tag{10}$$

The data provided by the US NCHS is a combination of reported deaths due to influenza and pneumonia. The pneumonia death rate is a combination of influenza viral pneumonia, secondary bacterial pneumonia that is a consequence of primary influenza infection and pneumonia due to causes other than influenza. This makes the estimation analysis a bit more complicated, because one should separate the dynamics of influenza from the pneumonia death rates before one can apply estimation procedure based on equation (8). However the dynamics of pneumonia are rather "slow" when compared to the influenza outbreaks. We could make a reasonable assumption that the reported death rate is $C(t) = D(t) + P(t)$, where $D(t)$ is the number of deaths per unit time due to influenza and $P(t)$ is the number of pneumonia deaths per unit time not caused by influenza. It is also natural to assume that $(dC(t)/dt) \gg (dP(t)/dt)$ so $P(t) \sim P_0 = constant$, meaning that the death rates due to pneumonia not due to influenza are fairly constant compared to the death rates due to the influenza. Taking this assumption one can write the equation for combined reported death rates $C(t)$ as follows:

$$\begin{array}{l} \dot{C} = (\xi_1 + 2\xi_2 P_0) C - \xi_2 \alpha_2 \int_0^t C dt C - \xi_2 C^2 - \\ (\xi_1 P_0 + \xi_2 P_0^2) + \xi_2 P_0 \alpha_2 (\int_0^t C dt + tC) - \xi_2 \alpha_2 P_0^2 t. \end{array} \tag{11}$$

It is convenient to rewrite equation (11) as

$$\dot{C} = \beta_1 C - \beta_2 \int_0^t C dt C - \beta_3 C^2 - \beta_4 + \beta_5 (\int_0^t C dt + tC) - \beta_6 t \tag{12}$$

where the parameters β_i are:

$$\begin{aligned}
\beta_1 &= (\xi_1 + 2\xi_2 P_0) \\
\beta_2 &= \xi_2 \alpha_2 \\
\beta_3 &= \xi_2 \\
\beta_4 &= (\xi_1 P_0 + \xi_2 P_0^2) \\
\beta_5 &= \xi_2 P_0 \alpha_2 \\
\beta_6 &= \xi_2 P_0^2 \alpha_2
\end{aligned} \tag{13}$$

4 Random Collocation Least Squares

Before providing a formal definition of RCLS , here is an informal description of how it works. Let us consider one differential equation for the variable $x(t)$ with a single constant coefficient α that needs to be determined,

$$\dot{x} = \alpha f(x(t), t) \tag{14}$$

In the integral form the equation above can be written

$$x(b) - x(a) = \alpha \int_a^b f(x(s), s) ds \tag{15}$$

For the sake of simplicity, assume that $f(x, t)$ is a linear function with respect to variable $x(t)$. If we have noiseless data for some interval $t \in [a, b]$ then for the non-singular case

$$\alpha = \frac{x(b) - x(a)}{\int_a^b f(x(s), s) ds} \tag{16}$$

If we have noisy data, i.e., $\tilde{x(t)} = x(t) + \epsilon(t)$, one can immediately see that the estimator (16) will be biased and interval dependent

$$\tilde{\alpha} = \frac{x(b) + \epsilon(b) - x(a) - \epsilon(a)}{\int_a^b f(x(s) + \epsilon(s), s) ds} \tag{17}$$

For linear $f(x, t)$ infinite data samples and zero-mean independently identically distributed noise (iid) ϵ

$$\int_a^b f(x(s) + \epsilon(s), s) ds = \lim_{\triangle s \to 0} \sum_i f(x_i + \epsilon_i, t_i) \triangle s \to \lim_{\triangle s \to 0} \sum_i f(x_i, t_i) \triangle s \tag{18}$$

Here the lim converges in mean-square (MS) and, therefore, also it converges in probability, see, e.g., [8]. To make the estimator (17) consistent,

we perform the following. If we generate on the interval $t \in [a,b]$ the smaller intervals $[t_{n_i}, t_{k_i}]$ (each of which we number as i), and each t_{n_i} and t_{k_i} are uniformly randomly distributed we get the following

$$\frac{1}{N}\sum_{i}^{N}(x(t_{k_i}) + \epsilon_{k_i} - x(t_{n_i}) - \epsilon_{n_i}) \longrightarrow \frac{1}{N}\sum_{i}^{N}(x(t_{k_i}) - x(t_{n_i})) \tag{19}$$

as $N \to \infty$. Since the equation (16) is valid for arbitratry intervals $[a,b]$ we can write

$$\frac{\sum_i^N(\tilde{x}(t_{k_i}) - \tilde{x}(t_{n_i}))}{\sum_i^N(\int_{t_{n_i}}^{t_{k_i}} f(\tilde{x}(s),s)ds} \to \alpha \tag{20}$$

as $N \to \infty$. From equation (20), we have a consistent estimator of α. We will apply now the same idea for estimation of parameters to the equation (11).

The data we work with has noise, which perhaps vaguely, we will associate with the accuracy of measurements. Let us rewrite equation (12) as follows:

$$\begin{array}{l} \dot{C} = \beta_1 C - \beta_2 \Psi C - \beta_3 C^2 - \beta_4 + \beta_5(\Psi + tC) - \beta_6 t \\ \dot{\Psi} = C, \Psi(0) = P_0 \end{array} \tag{21}$$

Now, introduce the following:

$$\begin{array}{l} \int_n^k \dot{C}dt = C_k - C_n \equiv \triangle_{k,n} \\ \int_n^k C dt \equiv I_{k,n} = I_{k,0} - I_{n,0} \\ \int_n^k C^2 dt \equiv J_{k,n} \\ \int_n^k \Psi C dt = \int_n^k \Psi\dot{\Psi}dt = (1/2)\Psi^2|_n^k = (1/2)I_{k,n}(I_{k,0} + I_{n,0}) \\ \int_n^k (\Psi + tC)dt = \int_n^k (\Psi + t\dot{\Psi})dt = \Psi t|_n^k = (I_{k,0}t_k - I_{n,0}t_n) \end{array} \tag{22}$$

Integration of first equation in (21) over the arbitrary time interval $[t_n, t_k]$ results in the following equation:

$$\begin{array}{l} \triangle_{k,n} = \beta_1 I_{k,n} - \beta_2(1/2)I_{k,n}(I_{k,0} + I_{n,0}) - \beta_3 J_{k,n} - \\ \beta_4(t_k - t_n) + \beta_5(I_{k,0}t_k - I_{n,0}t_n) - \beta_6(1/2)(t_k^2 - t_n^2) \end{array} \tag{23}$$

The measurements of death rates are $\tilde{C}_k = C(t_k) + \epsilon_k$, where ϵ_k is cumulative error due to measurement errors and fundamental model accuracy. We assume that the errors ϵ_k are independent and unbiased for each data point, i.e., $E[\epsilon] = 0$, where symbol $E[\ldots]$ means averaging over ensemble of measurement processes. Also assume that $E[\epsilon_k \epsilon_n] = \sigma^2 \delta_{kn}$, i.e., measurement errors have constant dispersion and represent additive noise. So far

we have not made any assumptions about the structure of the distribution of ϵ. For the measurements $\tilde{C}$ we have:

$$\begin{array}{l} \int_n^k \dot{\tilde{C}} dt = C_k - C_n + \epsilon_k - \epsilon_n = \triangle_{k,n} + \epsilon_{k,n} \equiv \tilde{\triangle}_{k,n} \\ \int_n^k \tilde{C} dt \simeq I_{k,n} \\ \int_n^k \tilde{C}^2 dt \simeq J_{k,n} + \sigma^2 (t_k - t_n) \\ \int_n^k \tilde{\Psi} \tilde{C} dt \simeq (1/2) I_{k,n} (I_{k,0} + I_{n,0}) \\ \int_n^k (\tilde{\Psi} + t\tilde{C}) dt \simeq (I_{k,0} t_k - I_{n,0} t_n) \end{array} \tag{24}$$

Since all ϵ_k are independent we have $E[\tilde{\triangle}_{k,n}] = \triangle_{k,n}$.

Let us now consider a set of integrations over the different time intervals $[t_{ni}, t_{ki}]$, i.e., we randomly sample the two points - the start of the interval t_{ni} and the end of the interval t_{ki}. For each sample interval $[t_{n_i}, t_{k_i}]$ we have a set of integrals $\tilde{\triangle}_{k_i,n_i}$, I_{k_i,n_i}, J_{k_i,n_i}, $I_{k_i,0}$ and $I_{n_i,0}$. We repeat this sampling process using a uniformly distributed random number generator to obtain the start and the end interval locations. Now, if we sum up all those linear relations together we get the following:

$$\begin{array}{c} \sum_i \tilde{\triangle}_{k_i,n_i} = \beta_1 \sum_i I_{k_i,n_i} - \beta_2 \frac{1}{2} \sum_i I_{k_i,n_i} (I_{k_i,0} + I_{n_i,0}) - \beta_3 \sum_i J_{k_i,n_i} - \\ (-\beta_3 \sigma^2 + \beta_4) \sum_i (t_{ki} - t_{ni}) + \beta_5 \sum_i (I_{k_i,0} t_{k_i} - I_{n_i,0} t_{n_i}) - \beta_6 (1/2) \sum_i (t_{k_i}^2 - t_{n_i}^2) \end{array} \tag{25}$$

We use the assumption here that in the limit of infinite data and infinite random sampling $\sum_i \tilde{\triangle}_{k_i,n_i} \equiv \sum_i \triangle_{k_i,n_i} + \epsilon_{k_i} - \epsilon_{n_i} = \sum_i \triangle_{k_i,n_i}$. For practical purposes we should have at least M sampling intervals, where M - is the number of data points.

We can now introduce the estimating function $F(\xi_1, \xi_2, P_0)$, where

$$\begin{array}{c} F = \sum_i \tilde{\triangle}_{k_i,n_i} - [(\xi_1 + 2\xi_2 P_0) \sum_i I_{k_i,n_i} - \xi_2 \alpha_2 \frac{1}{2} \sum_i I_{k_i,n_i} (I_{k_i,0} + I_{n_i,0}) - \\ \xi_2 \sum_i J_{k_i,n_i} - (-\xi_2 \sigma^2 + \xi_1 P_0 + \xi_2 P_0^2) \sum_i (t_{ki} - t_{ni}) + \\ \xi_2 \alpha_2 P_0 \sum_i (I_{k_i,0} t_{k_i} - I_{n_i,0} t_{n_i}) - \xi_2 \alpha_2 P_0^2 (1/2) \sum_i (t_{k_i}^2 - t_{n_i}^2)] \end{array} \tag{26}$$

The solution of the nonlinear problem $F \to 0$ provides the unknown parameters ξ_1, ξ_2 and P_0. One can see from (26) that for the fixed parameter P_0 the problem becomes linear in ξ_1 and ξ_2. Indeed we can rewrite the above equation as follows:

$$\begin{array}{c} F = \sum_i \tilde{\triangle}_{k_i,n_i} - \xi_1 [\sum_i I_{k_i,n_i} - P_0 \sum_i (t_{ki} - t_{ni})] + \\ \xi_2 [-\frac{1}{2} \alpha_2 \sum_i I_{k_i,n_i} (I_{k_i,0} + I_{n_i,0}) - \sum_i J_{k_i,n_i} + \alpha_2 P_0 \sum_i (I_{k_i,0} t_{k_i} - I_{n_i,0} t_{n_i}) + \\ 2P_0 \sum_i I_{k_i,n_i} - \alpha_2 P_0^2 (1/2) \sum_i (t_{k_i}^2 - t_{n_i}^2) - (P_0^2 - \sigma^2) \sum_i (t_{ki} - t_{ni})] \end{array} \tag{27}$$

The term $(P_0^2 - \sigma^2)$, where the noise σ^2 is unknown, can be reasonably well approximated by just P_0^2 if the noise level σ stays not greater then 10% of P_0.

To solve (27) we do a line search in P_0, i.e., we fix P_0 at various values, solve linear optimization problems $F(\xi_1, \xi_2 | P_0) \to min$, and then select ξ_1, ξ_2 and P_0 that provide the best minimum solution for F. Having solved (27) we estimate the parameters α_1 and α_3 using (10).

5 Results

As was mentioned above, we consider single-season outbreaks. The results of parameter estimation for the 1996-1997 and 1997-1998 seasons are presented in Figs. (2)-(7) and Tables 1 and 2. In the Figs. (2)-(3) we present a comparison of the data with the results of direct simulation of equation (12) with the parameters estimated using the same season data. One can confirm a rather adequate resolution of the epidemic amplitude, the location of the maximum, and the initial ramp. To estimate confidence intervals for the parameters we used parametric bootstrap (see Efron [7]). The results for the parameters are presented in Figs. (4) - (7). From the bootstrap histograms we estimate the mean, the median and 95% confidence intervals. Table 1 contains 1996-1997 season estimates for the parameters α, the basic reproductive number R_0, death level due pneumonia P_0, and the expectancy of a person infected by influenza person to die from the infection $\alpha_2/(\alpha_2 + \alpha_3)$. Table 2 contains the estimates for the season of 1997-1998. The parameter estimates were close for both epidemic seasons.

Using estimates for P_0 and α_3 for each of the seasons, we can now estimate the total number of infected people. Indeed, from the definition of the death rate $D(t) = \alpha_3 I(t)$, it follows that $I(t) = D(t)/\alpha_3$ or $I(t) = (C(t) - P_0)/\alpha_3$. The total number of infected people for one season then is

$$I_{tot} = \int I(t)dt = \int [C(t) - P_0]/\alpha_3 dt, \tag{28}$$

where integration is taken for one season. Our estimate of I_{tot} for the season of 1996-1997 is 44×10^6 and 49×10^6 for the season of 1997-1998. Taking into account that we used for our analysis the mortality data for 122 US metropolitan areas with total population about 130 million, the estimation for the whole country should double those numbers.

6 Discussion

In this paper, we developed a method for statistically separating influenza related mortality from other source of pneumonia mortality. Based on the SIR epidemic model, we estimated important epidemiological parameters. Our model provides good fits to the observed influenza mortality data for the two influenza seasons that we investigated. Thus, the model could be potentially used to predict influenza mortality in the future.
Our average estimate of the reproductive number is $R_0 = 1.3$. This indicates, that in a completely susceptible population, a randomly selected infected person would infect on the average 1.3 other people. This number is somewhat lower than previous estimates that ranged from 1.4 - 2.6 [6]. Further our mean estimate of the influenza case fatality ratio is 9.1×10^{-5}. This is somewhat less, but the same order of magnitude of previous estimates of around 33×10^{-5} using different, less direct methods [4] - [6]. Our average estimate of the total number of people infected in the U.S. for an influenza season is 93 million. This is somewhat larger than past estimates of around 60 million [6]. However, this latter estimate is based on reported influenza illness rates which are known to be underestimated. Thus, our model based estimates should prove useful in assessing influenza infection and fatality rates. We developed the method of random collocation least squares (RCLS). We compared the use of RCLS method with the Extended Kalman Filtering (EKF) applied to SIR and found the RCLS more robust and accurate mostly because linearized system (1) used by EKF is globally unstable along the solution trajectory. For small sets of highly noisy data we found that RCLS works better than many traditional estimation tools including regular least-squares. Part of the reason for this is because we work with differential equations and process time-derivative information which acts as a high-pass filter of noisy data. Recently, Markov chain Monte Carlo methods have been used with success for fitting SIR models to epidemic data [13]. We have modeled two seasons and found parameters for each of those seasons. The model and estimation procedure that we have developed here can be used in the future to help project the mortality rates for future epidemics. We can use these methods to estimate important epidemiologic parameters, such as R_0, from reported mortality data. In the past, a similar model has been coupled with the global transportation network to predict the spread of pandemic influenza [10] - [11]. Our plan is to adapt the model developed here to predict yearly spread of influenza among population centers in the US. This approach will be further adapted to predict the spread of other infectious agents that arise in nature or are intentionally released

as in the possible case of bioterrorist smallpox [14].

7 Acknowledgments

This work was performed under the auspices of the US Department of Energy and supported under the contract W-7405-ENG36. I. Longini's research was partially supported by NIH grant R01-AI32042.

References

[1] CDC. Prevention and control of influenza: Recommendations of the Advisory Committee on Immunization Practices (ACIP). MMWR 20 April; 50 (RR04): 1-46.

[2] Lui K-J, Kendal AP. Impact of influenza epidemics on mortality in the United States from October 1972 to May 1985. Am J Public Health 1987;77:712-6.

[3] Simonsen L, Schonberger LB, Stroup DF, Arden NH, Cox NJ. Impact of influenza on mortality in the USA. In: Brown LE, Hampson AW, Webster RG, eds. Options for the control of influenza III: proceedings of the 3rd International Conference on Options for the Control of Influenza, Cairns, Australia, 4–9 May, 1996. Amsterdam, Holland: Elsevier Science, 1996:26-33.

[4] Choi K, Thacker SB. An evaluation of influenza mortality surveillance, 1962-1979: II: Percentage of pneumonia and influenza as an indicator of influenza activity Am J Epidemiol 1981;113:227-35.

[5] L Simonsen, MJ Clarke, GD Williamson, DF Stroup, NH Arden, LB Schonberger. The impact of influenza epidemics on mortality: Introducing a severity index, Am J Public Health. 1997;87:1944-50.

[6] Weycker, D., Edelsberg, J., Halloran, M.E., Longini, I.M., Nizam A., Ciuryla, V. and Oster, G.: Clinical and economic benefits of routinely vaccinating young children against influenza. (Submitted).

[7] Efron, B., and Tibshirani, R. (1993) *An Introduction to the Bootstrap.* Chapman and Hall, New York.

[8] Papoulis, A. (1991) *Probability, Random Variables, and Stochastic Processes* WCB/McGraw-Hill.

[9] Longini, I.M.: The discrete-time general epidemic model: a synthesis. Mathematical Biosciences 81, 19-41 (1986).

[10] Rvachev, L.A. and Longini, I.M.: A mathematical model for the global spread of influenza. Mathematical Biosciences, 75:3-22 (1985).

[11] Longini, I.M.: A mathematical model for predicting the geographic spread of new infectious agents. Mathematical Biosciences 90, 367-383 (1988).

[12] Chin, J. Control of Communicable Diseases, 17th addition, America Public Health Assoc. (2000)

[13] O'Neill P.D., Balding D.J., Becker N.G., Eerola M. and Mollison D.: Analyses of infectious disease data from household outbreaks by Markov Chain Monte Carlo methods. Appl. Stat. 49: 517-542 (2000).

[14] Fauci, A.S.: Smallpox vaccination policy - the need for dialougue. New England Journal of Medicine, 346: 1319-20 (2002).

	mean	**median**	**95% conf.**
α_1	1.62	1.63	(1.59:1.67)
α_3	1.22×10^{-4}	1.23×10^{-4}	$(9.33 \times 10^{-5} : 1.54 \times 10^{-4})$
$R_0 = \alpha_1/\alpha_2$	1.24	1.26	(1.22:1.28)
$\alpha_3/(\alpha_2 + \alpha_3)$	9.41×10^{-5}	9.47×10^{-5}	$(7.20 \times 10^{-5} : 1.19 \times 10^{-4})$
P_0	658	662	(630:690)

Table 1: Parametric bootstrap estimates of parameters for the epidemic season of 1996-1997

	mean	**median**	**95% conf.**
α_1	1.69	1.70	(1.65:1.73)
α_3	1.13×10^{-4}	1.13×10^{-4}	$(8.80 \times 10^{-5} : 1.38 \times 10^{-4})$
$R_0 = \alpha_1/\alpha_2$	1.30	1.31	(1.27:1.34)
$\alpha_3/(\alpha_2 + \alpha_3)$	8.67×10^{-5}	8.73×10^{-5}	$(6.80 \times 10^{-5} : 1.07 \times 10^{-4})$
P_0	663	666	(630:702)

Table 2: Parametric bootstrap estimates of parameters for the epidemic season of 1997-1998

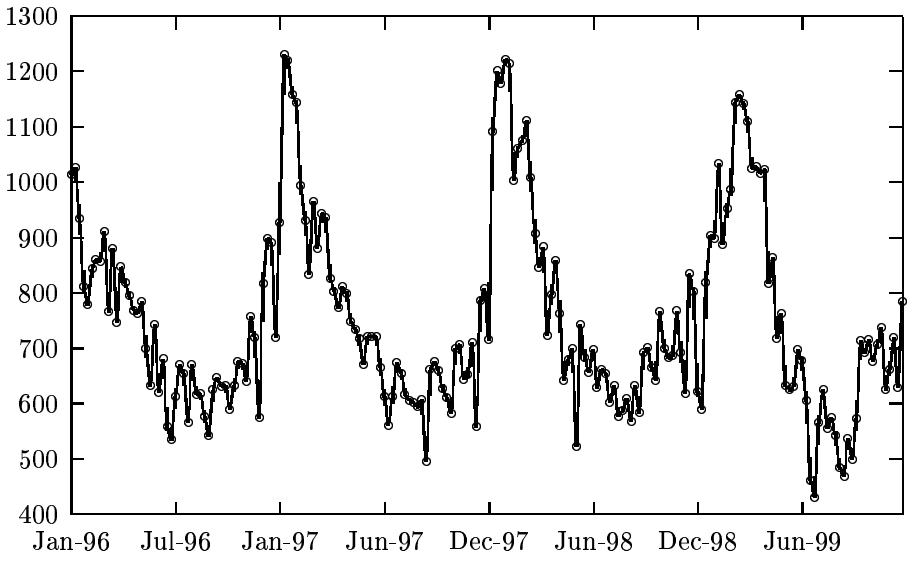

Figure 1: Mortality and Morbidity data for pneumonia and influenza. Excess death rates for all 122 US cities combined

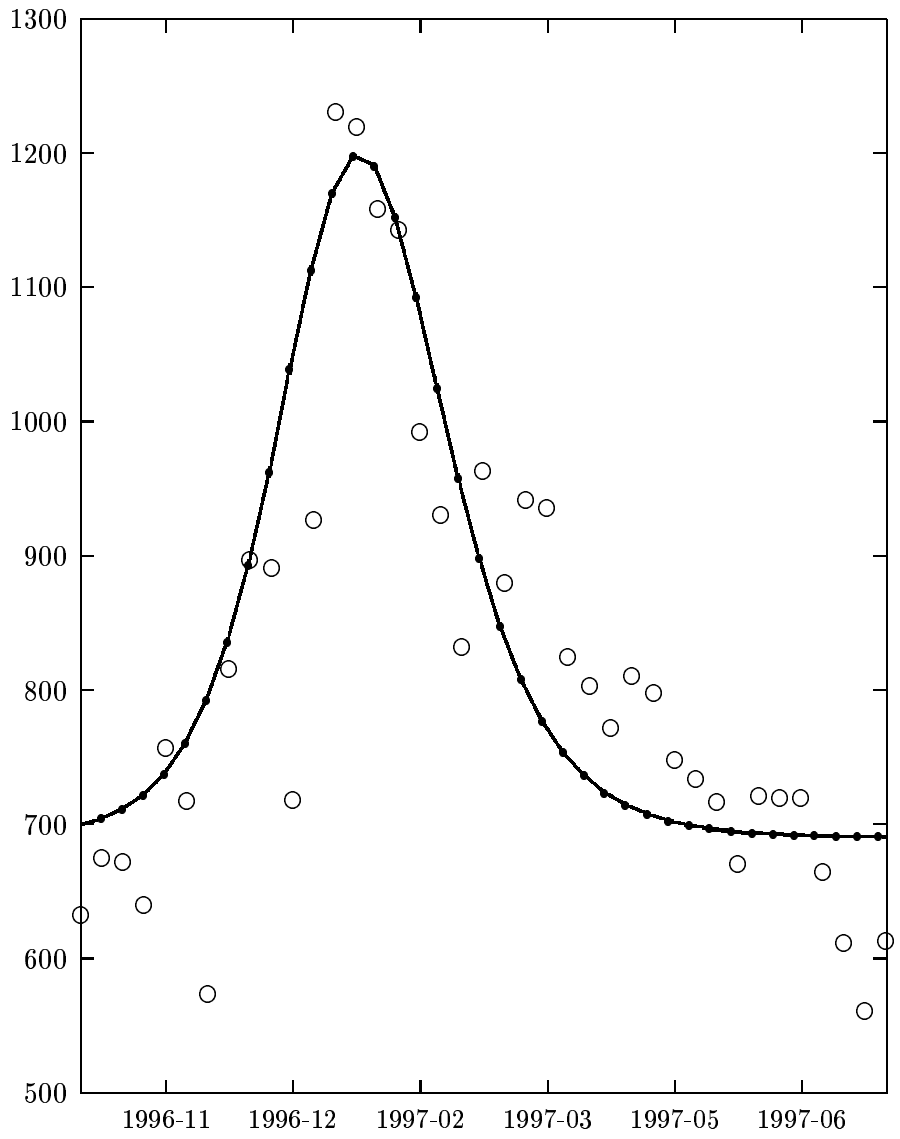

Figure 2: Comparison of actual data for total US reported death rates in 1996-1997 vs simulation with estimated parameters obtained by RCLS. Circles - weekly MMWR data, solid line - weekly sampled (integrated) solution of equation (11)

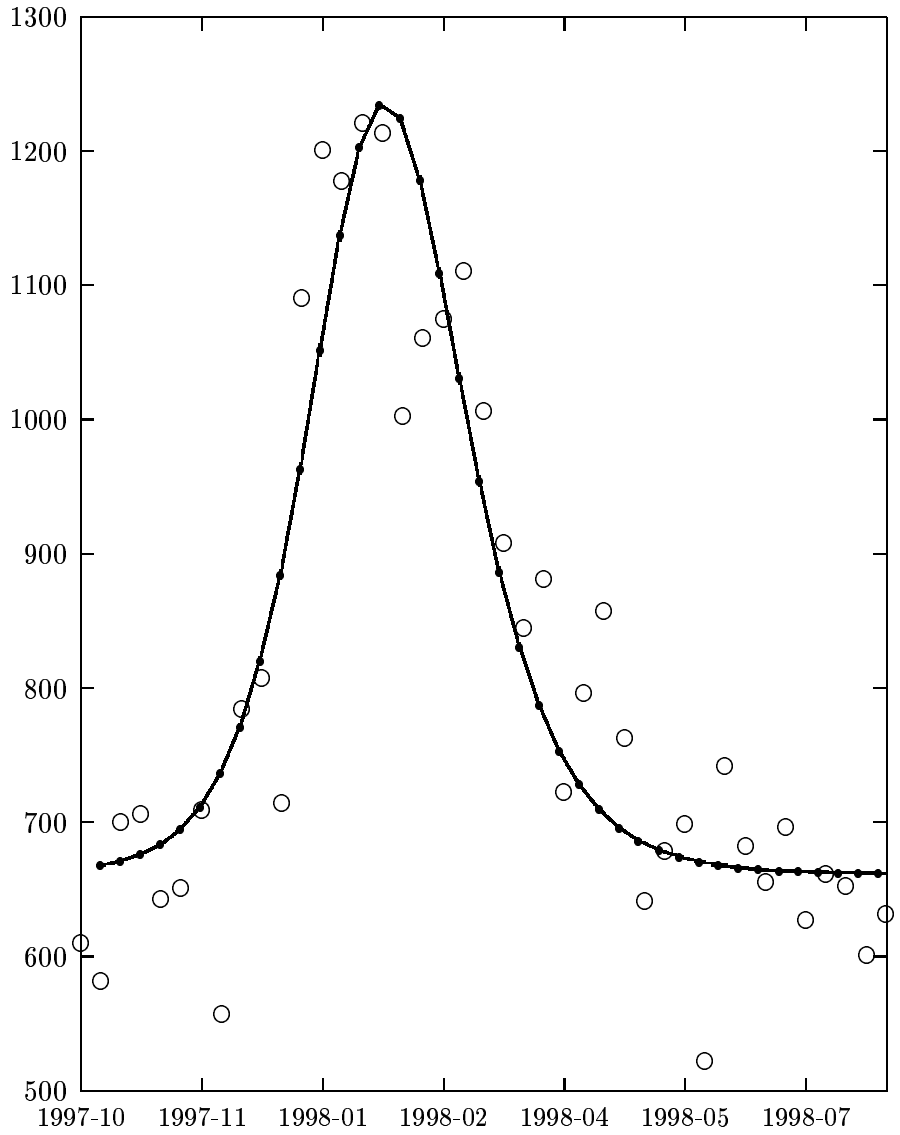

Figure 3: Comparison of actual data for total US reported death rates in 1997-1998 vs simulation with estimated parameters obtained by RCLS. Circles - weekly MMWR data, solid line - weekly sampled (integrated) solution of equation (11)

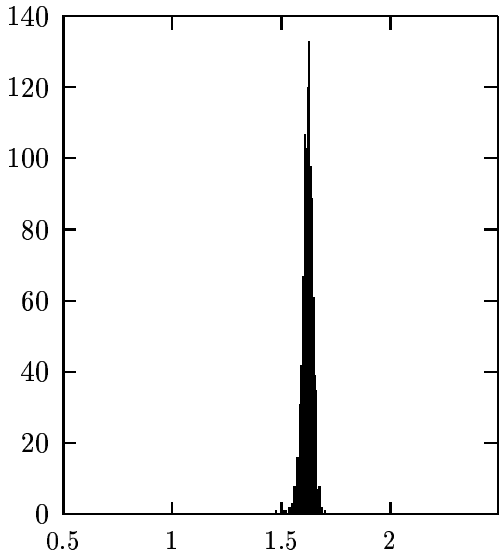

Figure 4: α_1

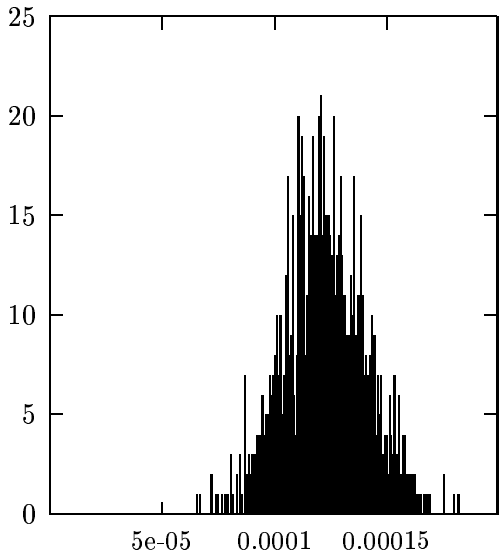

Figure 5: α_3

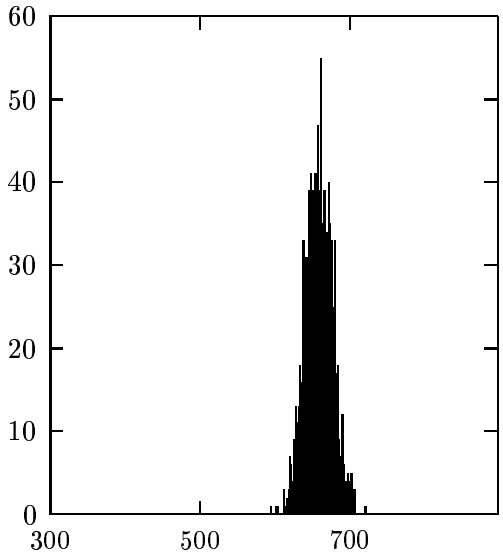

Figure 6: P_0

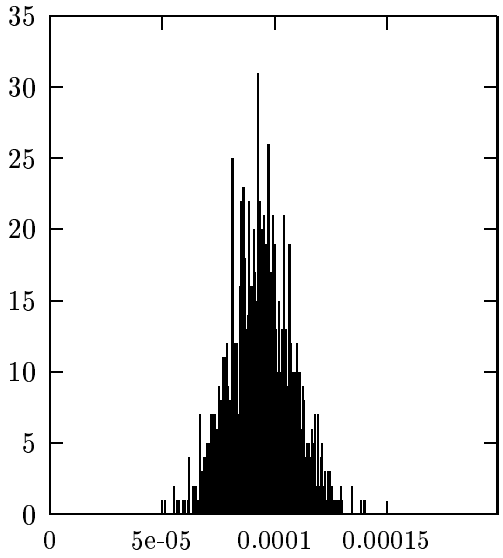

Figure 7: $\alpha_3/(\alpha_2 + \alpha_3)$